QUADRATURE FORMULAS FOR MONOTONE FUNCTIONS

ERICH NOVAK

Dedicated to Professor D. Kölzow on the occasion of his 60th birthday.

ABSTRACT. We prove that adaptive quadrature formulas for the class of monotone functions are much better than nonadaptive ones, if the average error is considered. Up to now it was only known that adaptive methods are not better in the worst case (for this and many other classes of functions) or in various average case settings.

We also prove that adaptive Monte Carlo methods are much better than nonadaptive ones. This also contrasts with analogous results for other classes (Sobolev classes, Hölder classes) where adaptive methods are only slightly better than nonadaptive ones.

1. Introduction and Historical Remarks

Our problem is to find a formula based solely on function evaluation at n points of $[0,1]$, which approximates $S(f)=\int_0^1 f(x)\,dx$ with small error over the class

$$F=\{f:[0,1]\to[0,1]\mid f \text{ monotone}, f(0)=0, f(1)=1\}.$$

This problem was studied in Kiefer [7], which is one of the first papers written in the spirit of "optimal recovery" or "information-based complexity". Kiefer considered all (nonrandomized) sequential (or adaptive) methods. He proved that the best method (with respect to the maximal error over the class F) is given by the trapezoidal rule

$$\widetilde{S}_n^T(f)=\frac{1}{2n+2}+\frac{1}{n+1}\sum_{i=1}^{n} f\left(\frac{i}{n+1}\right). \tag{1}$$

Hence we have an affine, nonlinear, and nonsequential algorithm which is optimal in the worst case. The maximal error of this $\widetilde{S}_n$ is $1/(2n+2)$.

We indicate some questions and comments.

a) Kiefer emphasizes in his paper that his method of proof is quite different from the method which is most commonly used in the literature of games and statistics.

1980 Mathematics Subject Classification (1985 Revision). Primary 41A55, 65D30, 65C05.

As a consequence, Kiefer's result is valid only for nonrandomized methods. I do not know of any paper studying randomized (or Monte Carlo) methods for this class F.

b) A similar result is also contained in Zwick [18], see Theorem 3 of this paper. Zwick (who only studied nonadaptive methods) notes that the error of the midpoint rule is $1/(2n)$. It is known (see Zubrzycki [17, Theorem 1]) that the latter formula is optimal within the class of linear methods.

c) Kiefer and Zubrzycki state their results on approximate integration of functions in the form of statements about estimation games between nature and the statistician. See Packel [10] for a similar diction.

d) The optimal method is an affine method. By a general theorem of Sukharev [12], affine methods are optimal within the class of all nonadaptive methods for every convex set F.

e) Observe that our problem is not a linear problem (in the usual sense of information-based complexity, see for example Traub et al. [15]), because the set F is not symmetric. For linear problems $S : F \to \mathbf{R}$ it is known, see Bakhvalov [2], that arbitrary adaptive methods are not better than linear nonadaptive methods.

f) It is known that randomized methods are much better than nonrandomized (or pure) methods for many classes of functions, see Novak [9]. In all known cases, however, adaptive Monte Carlo methods are only slightly better than nonadaptive ones. In this paper we will prove that adaptive Monte Carlo methods for the class F are much better than nonadaptive methods while nonadaptive Monte Carlo methods are only slightly better than (adaptive or nonadaptive) deterministic methods.

g) Suldin [14] was the first who studied average errors of quadrature formulas. He considered the classical Wiener measure and found the best method within the class of linear nonadaptive methods. Lee and Wasilkowski [8] studied the approximation of a linear functional on a Banach space with a Gaussian measure. It is known that adaptive methods are not better than nonadaptive ones for Gaussian and some other measures, see Traub et al. [15] for a survey. In Section 3 we study average errors with respect to the natural measure P of Graf et al. [4]. The average error $\Delta_P(\widetilde{S}_n)$ of optimal nonadaptive methods is of the order

$$\Delta_P(\widetilde{S}_n) \asymp n^{-\log 6/\log 4}.$$

This result is due to Graf, Novak [5]. We will prove that adaptive methods are much better—if we consider the average error with respect to P.

h) Sukharev introduced the concept of sequentially optimal algorithms as algorithms making in each step the best use (with respect to the worst-case information in the following steps) of the information given by $f(x_1), \ldots, f(x_i)$. A survey of this concept can be found in Sukharev [13]. This paper also contains a discussion of the numerical integration of monotone functions as an example. Numerical comparisons on stochastically generated monotone functions show that sequentially optimal algorithms are much better than $\widetilde{S}_n^T$, see Glinkin and Sukharev [3].

i) Wasilkowski and Gao [16] study the integration problem in a probabilistic setting for functions with a singular point. They show that an integration algorithm that adaptively locates the singular point is (probabilistically) superior to nonadaptive algorithms.

2. Monte Carlo methods

2.1 Nonadaptive methods. Assume that we want to approximate the functional $S : F \to \mathbf{R}$, given by $S(f) = \int f(x)\,dx$, by means of a mapping $\widetilde{S}_n(f) = \phi(f(x_1), \ldots, f(x_n))$ with some fixed knots $x_1 < x_2 < \ldots < x_n$ and an arbitrary $\phi : \mathbf{R}^n \to \mathbf{R}$. For convenience we also put $x_0 = 0$ and $x_{n+1} = 1$. It is easy to prove that the maximal error

$$\Delta_{\max}(\widetilde{S}_n) = \max_{f \in F} |S(f) - \widetilde{S}_n(f)|$$

of the optimal $\widetilde{S}_n$ using these knots $x_1, \ldots, x_n$ is given by

$$\Delta_{\max}(\widetilde{S}_n) = \max_{i=1,\ldots,n+1} |x_i - x_{i-1}|.$$

We see that equidistributed points yield the optimal information with the optimal $\widetilde{S}_n^T$, given by (1).

Now we consider (nonadaptive) Monte Carlo methods $\widetilde{S}_n^\omega$, given by $\widetilde{S}_n^\omega(f) = \phi_\omega(f(x_1^\omega), \ldots, f(x_n^\omega))$, where $\phi = \phi_\omega$ and the $x_i = x_i^\omega$ are random variables. We always assume that the functions $\omega \mapsto \widetilde{S}_n^\omega(f)$ are integrable. The error of such a randomized method is defined by

$$\Delta_{\max}(\widetilde{S}_n^\omega) = \max_{f \in F} E(|S(f) - \widetilde{S}_n^\omega(f)|),$$

where E means expectation of a random variable.

It is well known that for many classes of functions such nonadaptive Monte Carlo methods are much better than deterministic methods, i.e. the numbers $\inf_{\widetilde{S}_n^\omega} \Delta_{\max}(\widetilde{S}_n^\omega)$ are much smaller than the respective numbers $\inf_{\widetilde{S}_n} \Delta_{\max}(\widetilde{S}_n)$, see [9]. We stress that this result for Hölder spaces is due to Bakhvalov [1].

First we show that nonadaptive Monte Carlo methods for the class F are only slightly better than deterministic methods. We conclude that nonadaptive randomized methods are not interesting for this special class F.

Theorem 1.
For each nonadaptive Monte Carlo method using n function values the error bound

$$\Delta_{\max}(\widetilde{S}_n^\omega) \geq \frac{1}{8n}$$

holds.

Proof: We divide the interval $[0,1]$ in $2n$ sub-intervals of equal length and see that there is an interval $I \subset [0,1]$ of length $l = 1/(2n)$ such that there is no knot x_i^ω in I with probability at least $1/2$. Consider the functions $f_1, f_2 \in F$ which are both 0 to the left of I and 1 to its right, and $f_1 = 1$ and $f_2 = 0$ on I. We have $S(f_1) - S(f_2) = 1/(2n)$ and $\widetilde{S}_n^\omega(f_1) = \widetilde{S}_n^\omega(f_2)$ with probability at least $1/2$. The error of $\widetilde{S}_n^\omega$ can be estimated by

$$\Delta_{\max}(\widetilde{S}_n^\omega) \geq \max_{i=1,2} E(|S(f_i) - \widetilde{S}_n^\omega(f_i)|) \geq \frac{1}{4} \cdot |S(f_1) - S(f_2)| \geq \frac{1}{8n}.$$

2.2 Adaptive methods. Now we study adaptive or sequential methods, i.e. the knot x_i may depend on $f(x_1), \ldots, f(x_{i-1})$. The (nonadaptive) method $\widetilde{S}_n^T$, given by (1), is optimal even in the class of adaptive deterministic methods. This is due to Kiefer [7]. We shall see that adaptive randomized methods $\widetilde{S}_n^\omega$ are much better. To prove a lower bound in the spirit of Bakhvalov (see [1] or [9, 2.2.4]) we need the following lemma which should be known although I do not know a reference.

Lemma. For each natural number n we have

$$\sum_{k=0}^{n} \binom{n}{k} |2k - n| \geq 2^{n-1} \cdot \sqrt{2n}.$$

Proof: This inequality is true for $n = 1$ and $n = 2$. Now we assume that $n > 2$. We define

$$j_n = \sum_{k=0}^{n} \binom{n}{k} |2k - n| = 2 \cdot \sum_{k<n/2} \binom{n}{k} (n - 2k),$$

$a_n = \sum_{k<n/2} \binom{n}{k}$, and $b_n = \sum_{k<n/2} \binom{n-1}{k-1}$. By a simple calculation we obtain

$$\sum_{k<n/2} k \cdot \binom{n}{k} = n \cdot \sum_{k<n/2} \binom{n-1}{k-1}$$

and hence we have

$$j_n = 2na_n - 4nb_n.$$

Let n be even. Then we have $a_n = 2^{n-1} - \binom{n}{n/2}/2$ and $b_n = 2^{n-2} - \binom{n-1}{(n-1)/2}$ and obtain

$$j_n = 4n\binom{n-1}{(n-2)/2} - n\binom{n}{n/2} = 4(n-1)\binom{n-2}{(n-2)/2}.$$

Similarly we get

$$j_n = 2n\binom{n-1}{(n-1)/2}$$

if n is an odd number. Now we use the inequality

$$\binom{2k}{k} > \frac{4^k}{2\sqrt{k}},$$

true for all $k \in \mathbf{N}$ with $k > 1$, see Sierpiński [11, p. 132], and get

$$j_n > 4(n-1)\frac{2^{n-2}}{2\sqrt{(n-2)/2}} = \frac{\sqrt{2}(n-1)}{\sqrt{n-2}} \cdot 2^{n-1} > \sqrt{2n} \cdot 2^{n-1}$$

for even $n \geq 4$ and

$$j_n > 2n\frac{2^{n-1}}{2\sqrt{(n-1)/2}} > \sqrt{2n} \cdot 2^{n-1}$$

for odd $n \geq 3$.

Theorem 2.
For each adaptive Monte Carlo method using n function values the error bound

$$\Delta_{\max}(\widetilde{S}_n^\omega) \geq \frac{\sqrt{2}}{32} \cdot n^{-3/2}$$

holds.

Proof: For given $n \in \mathbf{N}$ we divide the interval $[0,1]$ in $2n$ small intervals of length $1/(2n)$. We consider the set F_0 of all step functions which are constant on each small intervall with $f(x) = x$ for one of the endpoints of each interval. The set F_0 consists of 2^{2n} functions and $F_0 \subset F$.

Now let μ be the equidistribution on the set F_0 and let $\widetilde{S}_n$ be any (adaptive or nonadaptive) deterministic method using n function values. Collecting those $f \in F_0$ with the same information we obtain the following estimate for the mean value of the error of $\widetilde{S}_n$ with respect to μ.

$$E_\mu(|S - \widetilde{S}_n|) = 2^{-2n} \sum_{f \in F_0} |S(f) - \widetilde{S}_n(f)| \geq 2^{-n} \sum_{i=0}^{n} \binom{n}{i} |i - n/2| \cdot \frac{\epsilon}{2}.$$

The number $\epsilon = (1/2n)^2$ is given by the difference $|S(f_1) - S(f_2)|$ for two functions $f_1, f_2 \in F_0$ which are different only in one of the small intervals. Using our Lemma we get

$$E_\mu(|S - \widetilde{S}_n|) \geq \frac{\sqrt{2}}{32} n^{-3/2}.$$

Now the statement of Theorem 2 follows by the theorem of Fubini, i.e.

$$\Delta_{\max}(\widetilde{S}_n^\omega) \geq \int_F E(|S(f) - \widetilde{S}_n^\omega(f)|) d\mu(f) = E\left(\int_F |S(f) - \widetilde{S}_n^\omega(f)| d\mu(f)\right) \geq \frac{\sqrt{2}}{32} \cdot n^{-3/2}.$$

Remark: The known Monte Carlo integration methods, as constructed by Bakhvalov [1], Haber [6], Novak [9] and many others, are nonadaptive. For many classes F of functions this does not mean a serious restriction because it can be shown that adaptive Monte Carlo methods are only slightly better than nonadaptive ones. This result is due to Bakhvalov [1] for Hölder classes and can in the same way be proved for Sobolev classes, see [9].

The situation is different for our present class F of monotone functions. We want to show that adaptive methods are much better than nonadaptive ones. More exactly, we will prove that the lower bound of Theorem 2 is optimal with respect to the order $n^{-3/2}$ of convergence. To prove this result we have to construct a new type of stochastic quadrature formulas. We try to give a construction which is as simple as possible:

First step: Let $m \in \mathbf{N}$. We calculate the function values $f(a_i)$, $i = 1, \ldots, m$, for the fixed knots $a_i = i/(m+1)$. The choice of these knots is nonadaptive and deterministic. We also put $a_0 = 0$ and $a_{m+1} = 1$.

Second step: For each sub-interval $[a_i, a_{i+1}]$ we compute the respective area

$$F_i = (a_{i+1} - a_i) \cdot (f(a_{i+1}) - f(a_i)).$$

In the case $F_i \leq 1/(m+1)^2$ the respective interval will not be divided further. In the case $1/(m+1)^2 < F_i \leq 2/(m+1)^2$ the interval $[a_i, a_{i+1}]$ will be divided in two intervals of the same length, and so on: In the case $k/(m+1)^2 < F_i \leq (k+1)/(m+1)^2$ the interval $[a_i, a_{i+1}]$ will be divided in $(k+1)$ intervals of the same length. In this way we divide some of the $[a_i, a_{i+1}]$ and obtain altogether $r(f)$ intervals. Observe that $r(f) \leq 2m+1$ for every $f \in F$. The area (as defined above) of each of the new intervals I_k is at most $1/(m+1)^2$. The construction of the I_k clearly is an adaptive (but non-stochastic) step in our construction.

Third step: Now we take $b_k \in I_k$, independent and equidistributed. We estimate the integral of f by the formula

$$\widetilde{S}_n^\omega(f) = \sum_{i=1}^{r(f)} f(b_k(\omega)) \cdot \lambda(I_k)$$

and observe that the number of used function values is $r(f)+m \le 3m+1$. Hence we can define $n = 3m+1$. Now it is easy to prove the following theorem.

Theorem 3.
The method $\widetilde{S}^{\omega}_{3m+1}$ defined above satisfies

$$\Delta_{\max}(\widetilde{S}^{\omega}_{3m+1}) \le \tfrac{1}{2}\sqrt{2}\cdot(m+1)^{-3/2}$$

and is optimal with respect to the order of convergence.

Proof: The random variables $f(b_k(\omega))$ are independent and it is easy to prove that their variance $\sigma^2(f(b_k(\omega)))$ is bounded by $(\lambda(f(I_k)))^2/4$. Therefore we obtain the inequality

$$\sigma^2(\widetilde{S}^{\omega}_n(f)) \le \sum_{i=1}^{2m+1} \lambda(I_k)^2 \cdot \lambda(f(I_k))^2/4 \le \tfrac{1}{4}(2m+1)/(m+1)^4 < (m+1)^{-3}/2.$$

We conclude that

$$\Delta_{\max}(\widetilde{S}^{\omega}_n(f)) \le \sup_{f\in F} \sigma^2(\widetilde{S}^{\omega}_n(f))^{1/2} < \tfrac{1}{2}\sqrt{2}\cdot(m+1)^{-3/2}.$$

Remark: The upper bound of Theorem 3 can be written as $\Delta_{\max}(\widetilde{S}^{\omega}_n) \le \frac{1}{2}\sqrt{2}((n+2)/3)^{-3/2} < \frac{1}{2}\sqrt{54}n^{-3/2}$. Hence this bound differs by a factor of about 83 from the lower bound of Theorem 2.

Though it is easy to improve the upper bound slightly, it seems to be difficult to obtain the best constant in the inequality $\Delta_{\max}(\widetilde{S}^{\omega}_n) \le c\cdot n^{-3/2}$. I conjecture that the following method is not far from being optimal for arbitrary $n \in \mathbf{N}$:

Assume that the knots $x_1,\dots,x_k$ are already chosen, $k=0,\dots,n-1$. By renumbering the x_i we obtain $0 = b_0 \le b_1 \le \dots \le b_k \le b_{k+1} = 1$ with $x_i = b_j$ for some j. Now we compute k^* such that the number

$$(b_{k+1}-b_k)\cdot(f(b_{k+1})-f(b_k))$$

is maximal. We take $x_{k+1}(\omega)$ randomly from $[b_{k^*}, b_{k^*+1}]$, i.e. the random variable x_{k+1} is equidistributed in this interval. Having defined in this way the knots $x_1,\dots,x_n$ we approximate the integral $\int f(x)\,dx$ by the integral of the piecewise linear function g with $g(x_i) = f(x_i)$, $g(0) = 0$, $g(1) = 1$. It is not easy, however, to obtain nice error estimates for this Monte Carlo method.

One also would expect that the upper bound in Theorem 3 could be improved by using all $r(f)+m$ values of the integrand in approximating the integral, and furthermore, by using the trapezoidal rule instead of the coarser rectangular rule. Again it is difficult, however, to obtain good error estimates for this method.

3. Average case results

3.1 Nonadaptive methods. The purpose of this section is to discuss the average error of (deterministic) quadrature formulas $\widetilde{S}_n$. If one wants to consider average errors then one of the problems is to define a suitable probability measure on the function set in question. We use the "natural" measure P on F which was studied by Graf et al. [4].

Let F be equipped with the topology of uniform convergence. A typical P-random function $f \in F$ can be generated by the following construction. First choose the value $f(1/2)$ according to the uniform distribution on $[0,1]$. Then choose $f(1/4)$ according to the uniform distribution on $[0, f(1/2)]$ and independently $f(3/4)$ according to the uniform distribution on $[f(1/2), 1]$. Continue this process. With probability one, the function constructed in that way on the dyadic rationals extends to a (unique and continuous) element $f \in F$.

The average error $\Delta_P(\widetilde{S}_n)$ of a method $\widetilde{S}_n$ is defined by

$$\Delta_P(\widetilde{S}_n) = \left(\int_F (S(f) - \widetilde{S}_n(f))^2 \, dP(f) \right)^{1/2}.$$

Nonadaptive methods were studied in [5]. We summarize some of the results of that paper. It can be shown that $\Delta_P(\widetilde{S}^T_{2^m-1})^2 = 40^{-1/2} \cdot 6^{-m/2}$. It can also be proved that this gives the best possible rate of convergence for nonadaptive methods. As a consequence we obtain the following result.

Theorem 4.
For nonadaptive (deterministic) methods using n function values the error bound

$$\inf_{\widetilde{S}_n} \Delta_P(\widetilde{S}_n) \asymp n^{-\log 6/\log 4},$$

holds.

3.2 Adaptive methods. We want to show that adaptive methods are much better (on the average) than nonadaptive ones. This result contrasts with known results for Gaussian (and some other) measures where adaption does not help. Though we do not obtain the optimal order of convergence for adaptive methods, we explicitly construct a method $\widetilde{S}^a_n$ which should not be far from being optimal and prove the error bound $\Delta_P(\widetilde{S}^a_n) \le c \cdot n^{-3/2}$. Because of $\log 6/\log 4 \approx 1.2925$ this adaptive method is much better than nonadaptive methods.

Assume that the knots $x_1, \ldots, x_k$ are already chosen, $k = 0, \ldots, n-1$. By renumbering the x_i we obtain $0 = b_0 \le b_1 \le \ldots \le b_k \le b_{k+1} = 1$ with $x_i = b_j$ for some j. Now we compute k^* such that the number

$$(b_{k+1} - b_k) \cdot (f(b_{k+1}) - f(b_k))$$

is maximal. We take $x_{k+1} = (b_{k^*+1} - b_{k^*})/2$. Having defined in this way the knots $x_1, \ldots, x_n$ we approximate the integral $\int f(x)\,dx$ by the integral of the piecewise linear function g with $g(x_i) = f(x_i)$, $g(0) = 0$, $g(1) = 1$. We denote the method just defined by $\widetilde{S}_n^a$. The next theorem shows that it converges at least as $n^{-3/2}$. It would be interesting to have a better estimate for this special method and lower bounds for the average error of arbitrary adaptive methods. Some rather rough computer experiments show that $\Delta_P(\widetilde{S}_n^a) \le c \cdot n^{-1.63}$ and the exponent seems to be optimal.

Theorem 5.

$$\Delta_P(\widetilde{S}_n^a) \le 5 \cdot n^{-3/2}.$$

Proof: For each special $f \in F$ the method $\widetilde{S}_n^a$ uses some knots $0 = b_0 \le b_1 \le \ldots \le b_n \le b_{n+1} = 1$ with the respective function values $y_i = f(b_i)$. First we want to give a formula for the conditional average error of $\widetilde{S}_n^a$ under the condition $g(b_i) = y_i$.

We know from [5] that

$$E\left(\int f(x)\,dx\right) = 1/2 \qquad \text{and} \qquad E\left(\left(\int (f(x)\,dx - 1/2\right)^2\right) = 1/40,$$

where E means expectation with respect to P. The measure P is scaling invariant between $i2^{-m}$ and $(i+1)2^{-m}$ for $0 \le i < 2^m$ and $m \in \mathbf{N}$, see [4, Theorem 4.6]. Observe that our construction of the knots b_j implies that each interval $[b_j, b_{j+1}]$ is of the form $[i2^{-m}, (i+1)2^{-m}]$. Therefore given $g(b_i) = y_i$, $i = 1, \ldots, n$, the conditional mean value of

$$\left(\int_{b_i}^{b_{i+1}} g(x)\,dx - (b_{i+1} - b_i)(g(b_{i+1}) + g(b_i))/2\right)^2$$

is given by $\sigma_i^2 = (b_{i+1} - b_i)^2 (f(b_{i+1}) - f(b_i))^2/40$ and the random variables

$$\int_{b_i}^{b_{i+1}} g(x)\,dx - (b_{i+1} - b_i)(g(b_{i+1}) + g(b_i))/2$$

are conditionally independent. Hence we obtain the formula

$$\begin{aligned} \sigma^2(f) :=& E\left(\left(\int g(x)\,dx - \widetilde{S}_n^a(g)\right)^2 \mid g(b_i) = y_i,\ i = 1, \ldots, n\right) \\ =& \sum_{i=0}^{n} (b_{i+1} - b_i)^2 (y_{i+1} - y_i)^2/40. \end{aligned} \tag{2}$$

The expectation of this value gives an exact formula for the value of $\Delta_P(\widetilde{S}_n^a)^2$. The exact calculation of this integral seems to be difficult, however, and we only give a rough estimate showing that

$$\sigma^2(f) \le 25n^{-3}$$

for each $f \in F$ and therefore $\Delta_P(\widetilde{S}_n^a)^2 \le 25n^{-3}$.

For that we first define a different method $\widetilde{S}_n^*$ which is slightly worse than $\widetilde{S}_n^a$ but easier to analyze. Let $k \in \mathbf{N}$ and let $m+1 = 2^k$. We give a construction which is similar as that for the proof of Theorem 3.

First step: Compute $f(a_i)$ for $i = 1, \ldots, m$, where $a_i = i/(m+1)$. Again we put $a_0 = 0$ and $a_{m+1} = 1$.

Second step: For each sub-interval $[a_i, a_{i+1}]$ we consider the respective area

$$F_i = (a_{i+1} - a_i) \cdot (f(a_{i+1}) - f(a_i)).$$

In the case $F_i \le 1/(m+1)^2$ the respective interval will not be divided further. In the case $1/(m+1)^2 < F_i \le 2/(m+1)^2$ the interval $[a_i, a_{i+1}]$ will be divided in two intervals of the same length, and so on: In the case $2^l/(m+1)^2 < F_i \le 2^{l+1}/(m+1)^2$ the interval $[a_i, a_{i+1}]$ will be divided in 2^{l+1} intervals of the same length. In this way we divide some of the $[a_i, a_{i+1}]$ and obtain altogether at most $(3m+1)$ intervals. The area (as defined above) of each of the new intervals I_j^* is at most $1/(m+1)^2$. The construction of the I_j^* clearly is an adaptive step in our construction.

We define $n = 3m$ and consider the respective intervals $I_1^a, \ldots, I_n^a$ for the method $\widetilde{S}_n^a$, i.e., $I_k^a = [b_{k-1}, b_k]$. It follows from the construction of the b_k that we also have

$$(f(b_k) - f(b_{k-1})) \cdot (b_k - b_{k-1}) \ge \frac{1}{(m+1)^2}$$

for all k. Therefore the following estimate for $\widetilde{S}_n^*$ is also valid for $\widetilde{S}_n^a$.

Third step: Now we take each end of an interval I_j^* as a knot of the method $\widetilde{S}_n^*$. Observe that we have (at most) $n = 3m$ knots $x_1, \ldots, x_n$ together with $x_0 = 0$ and $x_{n+1} = 1$. We estimate the integral of f by the the integral of the piecewise linear function g with $g(x_i) = f(x_i)$, $i = 1, \ldots, n+1$.

Because of the special scaling invariance of the measure P we obtain

$$\sigma^2(f) \le (3m+1)(m+1)^{-4}/40,$$

an estimate which is analog to (2). Hence we have proved the error bound

$$\Delta_P(\widetilde{S}_n^*)^2 \le \frac{81(n+1)}{40(n+3)^4}$$

for $n = 3m = 3 \cdot 2^k - 3$ and we conclude that this bound also holds for $\widetilde{S}_n^a$. Therefore the estimate

$$\Delta_P(\widetilde{S}_n^a) \le 5 \cdot n^{-3/2}$$

holds for any $n \in \mathbf{N}$.

Remark: Assume now that $\widetilde{P}$ is an arbitrary (Borel) probability measure on F. Using the proof of Theorem 3, we easily get the bound

$$\inf_{\widetilde{S}_{3m+1}} \Delta_{\widetilde{P}}(\widetilde{S}_{3m+1}) \le \frac{\sqrt{2}}{2} \cdot (m+1)^{-3/2},$$

where the infimum is taken over all adaptive methods using $n = 3m + 1$ knots. This bound, however, is an abstract bound and it may be difficult to construct a suitable method $\widetilde{S}_n$, as we have done for P.

References

1. Bakhvalov, N. S., On approximate computation of integrals, Vestnik MGV, Ser. Math. Mech. Astron. Phys. Chem. **4** (1959), 3–18. [Russian]
2. Bakhvalov, N. S., On the optimality of linear methods for operator approximation in convex classes of functions, USSR Comput. Math. Math. Phys. **11** (1971), 244–249.
3. Glinkin, I. A., and A. G. Sukharev, Efficiency analysis of some algorithms of numerical integration and their applications to the solution of extremal problems. In: Voprosy kibernetiki. Akad. Nauk SSSR, Moscow **122** (1985), Models and Methods of Global Optimization (V. V. Fedorov, ed.), 23–37. [Russian]
4. Graf, S., R. D. Mauldin, and S. C. Williams, Random homeomorphisms, Advances in Math. **60** (1986), 239–359.
5. Graf, S., and E. Novak, The average error of quadrature formulas for functions of bounded variation, Rocky Mt. J. Math. to appear.
6. Haber, S., Numerical evaluation of multiple integrals, SIAM Review **12** (1970), 481–526.
7. Kiefer, J., Optimum sequential search and approximation methods under minimum regularity assumptions, J. Soc. Indust. Appl. Math. **5** (1957), 105–136.
8. Lee, D., and G. W. Wasilkowski, Approximation of linear functionals on a Banach space with a Gaussian measure, J. Complexity **2** (1986), 12–43.
9. Novak, E., *Deterministic and Stochastic Error Bounds in Numerical Analysis*, Lecture Notes in Math. **1349** (1988).
10. Packel, E. W., The algorithm designer versus nature: a game-theoretic approach to information-based complexity, J. Complexity **3** (1987), 244–257.
11. Sierpiński, W., *Elementary theory of numbers*, Warszawa, 1964.

12. Sukharev, A. G., On the existence of optimal affine methods for approximating linear functionals, J. Complexity **2** (1986), 317–322.
13. Sukharev, A. G., The concept of sequential optimality for problems in numerical analysis, J. Complexity **3** (1987), 347–357.
14. Suldin, A. V., Wiener measure and its application to approximation methods I, II, Izv. Vyssh. Uchebn. Zaved. Math. **13, 18**, 145–158, 165–179.
15. Traub, J. F., G. W. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity*, Academic Press, 1988.
16. Wasilkowski, G. W., and F. Gao, On the power of adaptive information for functions with singularities, Technical Report, University of Kentucky, 1989.
17. Zubrzycki, S., Some approximate integration formulas of statistical interest, Coll. Math. **11** (1963), 123–136.
18. Zwick, D., Optimal quadrature for convex functions and generalizations, ISNM **85** (1988), 310–315.

Mathematisches Institut, Universität Erlangen, Bismarckstr. 1 1/2, D-8520 Erlangen, FRG.